Memory Requirements in a Telephone Exchange

By CLAUDE E. SHANNON

(*Manuscript Received Dec. 7, 1949*)

1. Introduction

A GENERAL telephone exchange with N subscribers is indicated schematically in Fig. 1. The basic function of an exchange is that of setting up a connection between any pair of subscribers. In operation the exchange must "remember," in some form, which subscribers are connected together until the corresponding calls are completed. This requires a certain amount of internal memory, depending on the number of subscribers, the maximum calling rate, etc. A number of relations will be derived based on these considerations which give the minimum possible number of relays, crossbar switches or other elements necessary to perform this memory function. Comparison of any proposed design with the minimum requirements obtained from the relations gives a measure of the efficiency in memory utilization of the design.

Memory in a physical system is represented by the existence of stable internal states of the system. A relay can be supplied with a holding connection so that the armature will stay in either the operated or unoperated positions indefinitely, depending on its initial position. It has, then, two stable states. A set of N relays has 2^N possible sets of positions for the armatures and can be connected in such a way that these are all stable. The total number of states might be used as a measure of the memory in a system, but it is more convenient to work with the logarithm of this number. The chief reason for this is that the amount of memory is then proportional to the number of elements involved. With N relays the amount of memory is then $M = \log 2^N = N \log 2$. If the logarithmic base is two, then $\log_2 2 = 1$ and $M = N$. The resulting units may be called binary digits, or more shortly, bits. A device with M bits of memory can retain M different "yes's" or "no's" or M different 0's or 1's. The logarithmic base 10 is also useful in some cases. The resulting units of memory will then be called decimal digits. A relay has a memory capacity of .301 decimal digits. A 10×10 crossbar switch has 100 points. If each of these points could be operated independently of the others, the total memory capacity would be 100 bits or 30.1 decimal digits. As ordinarily used, however, only one point in a vertical can be closed. With this restriction the capacity is one decimal digit for each vertical, or a total of ten decimal digits. The panels used in a panel type exchange are another form of memory device. If the commutator in a panel has 500 possible levels, it has a memory capacity of log 500;8.97 bits or 2.7 decimal digits. Finally, in a step-by-step system, 100-point selector switches are used. These have a memory of two decimal digits.

Frequently the actual available memory in a group of relays or other devices is less than the sum of the individual memories because of artificial restrictions on the available states. For technical reasons, certain states are made inaccessible—if relay A is operated relay B must be unoperated, etc. In a crossbar it is not desirable to have more than nine points in the same horizontal operated because of the spring loading on the crossarm. Constraints of this type reduce the memory per element and imply that more than the minimum requirements to be derived will be necessary.

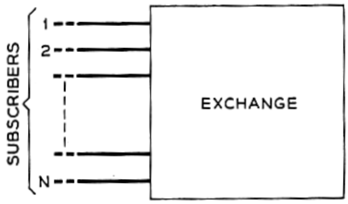

Fig. 1—General telephone exchange.

2. Memory Required for any S Calls out of N Subscribers

The simplest case occurs if we assume an isolated exchange (no trunks to other exchanges) and suppose it should be able to accommodate any possible set of S or fewer calls between pairs of subscribers. If there are a total of N subscribers, the number of ways we can select m pairs is given by

$$\frac{N(N-1)(N-2)\cdots(N-2m+1)}{2^m m!} = \frac{N!}{2^m m!(N-2m)!} \tag{1}$$

The numerator $N(N-1)\cdots(N-2m+1)$ is the number of ways of choosing the $2m$ subscribers involved out of the N. The $m!$ takes care of the permutations in order of the calls and 2^m the inversions of subscribers in pairs. The total number of possibilities is then the sum of this for $m = 0, 1, \cdots, S$; i.e.

$$\sum_{m=0}^{S} \frac{N!}{2^m m!(N-2m)!} \tag{2}$$

The exchange must have a stable internal state corresponding to each of these possibilities and must have, therefore, a memory capacity M where

$$M = \log \sum_{0}^{S} \frac{N!}{2^m m!(N-2m)!}. \tag{3}$$

If the exchange were constructed using only relays it must contain at least $\log_2 \sum N!/2^m m!(N - 2m)!$ relays. If 10×10 point crossbars are used in the normal fashion it must contain at least $\frac{1}{10} \log_{10} \sum N!/2^m m!(N - 2m)!$ of these, etc. If fewer are used there are not enough stable configurations of connections available to distinguish all the possible desired interconnections. With $N = 10{,}000$, and a peak load of say 1000 simultaneous conversations $M = 16{,}637$ bits, and at least this many relays or 502 10×10 crossbars would be necessary. Incidentally, for numbers N and S of this magnitude only the term $m = S$ is significant in (3).

The memory computed above is that required only for the basic function of remembering who is talking to whom until the conversation is completed. Supervision and control functions have been ignored. One particular supervisory function is easily taken into account. The call should be charged to

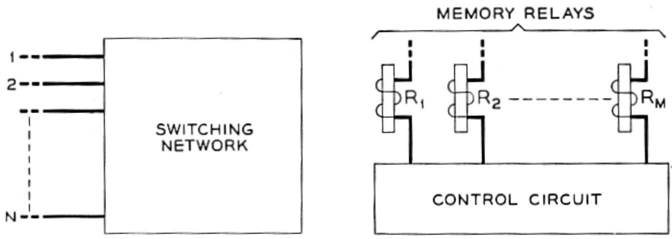

Fig. 2—Minimum memory exchange.

the calling party and under his control (i.e. the connection is broken when the calling party hangs up). Thus the exchange must distinguish between a calling b and b calling a. Rather than count the number of pairs possible we should count the number of ordered pairs. The effect of this is merely to eliminate the 2^m in the above formulas.

The question arises as to whether these limits are the best possible—could we design an exchange using only this minimal number of relays, for example? The answer is that such a design is possible in principle, but for various reasons quite impractical with ordinary types of relays or switching elements. Figure 2 indicates schematically such an exchange. There are M memory relays numbered $1, 2, \ldots, M$. Each possible configuration of calls is given a binary number from 0 to 2^M and associated with the corresponding configuration of the relay positions. We have just enough such positions to accommodate all desired interconnections of subscribers.

The switching network is a network of contacts on the memory relays such that when they are in a particular position the correct lines are connected together according to the correspondence decided upon. The control circuit is essentially merely a function table and requires, therefore, no memory. When a call is completed or a new call originated the desired configuration of the holding relays is compared with the present configuration and voltages applied to or eliminated from all relays that should be changed.

Needless to say, an exchange of this type, although using the minimum memory, has many disadvantages, as often occurs when we minimize a design for one parameter without regard to other important characteristics. In particular in Fig. 2 the following may be noted: (1) Each of the memory relays must carry an enormous number of contacts. (2) At each new call or completion of an old call a large fraction of the memory relays must change position, resulting in short relay life and interfering transients in the conversations. (3) Failure of one of the memory relays would put the exchange completely out of commission.

3. The Separate Memory Condition

The impracticality of an exchange with the absolute minimum memory suggests that we investigate the memory requirements with more realistic assumptions. In particular, let us assume that in operation a separate part of the memory can be assigned to each call in progress. The completion of a current call or the origination of a new call will not disturb the state of the memory elements associated with any call in progress. This assumption is reasonably well satisfied by standard types of exchanges, and is very natural to avoid the difficulties (2) and (3) occurring in an absolute minimal design.

If the exchange is to accommodate S simultaneous conversations there must be at least S separate memories. Furthermore, if there are only this number, each[1] of these must have a capacity $\log \frac{N(N-1)}{2}$. To see this, suppose all other calls are completed except the one in a particular memory. The state of the entire exchange is then specified by the state of this particular memory. The call registered here can be between any pair of the N subscribers, giving a total of $N(N-1)/2$ possibilities. Each of these must correspond to a different state of the particular memory under consideration, and hence it has a capacity of least $\log N(N-1)/2$.

The total memory required is then

$$M = S \log \frac{N(N-1)}{2}. \tag{4}$$

If the exchange must remember which subscriber of a pair originated the call we obtain

$$M = S \log N(N-1). \tag{5}$$

or, very closely when N is large,

$$M = 2S \log N. \tag{6}$$

[1] B. D. Holbrook has pointed out that by using more than S memories, each can have for certain ratios of $\frac{S}{N}$, a smaller memory, resulting in a net saving. This only occurs, however, with unrealistically high calling rates.

The approximation in replacing (5) by (6), of the order of $\frac{S}{N} \log e$, is equivalent to the memory required to allow connections to be set up from a subscriber to himself. With $N = 10{,}000$, $S = 1{,}000$, we obtain $M = 26{,}600$

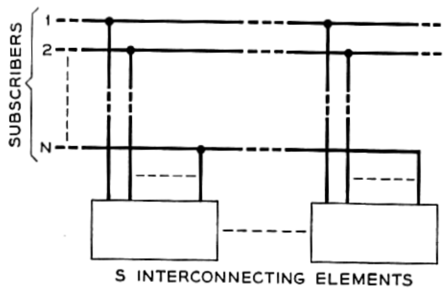

Fig. 3—Minimum separate memory exchange.

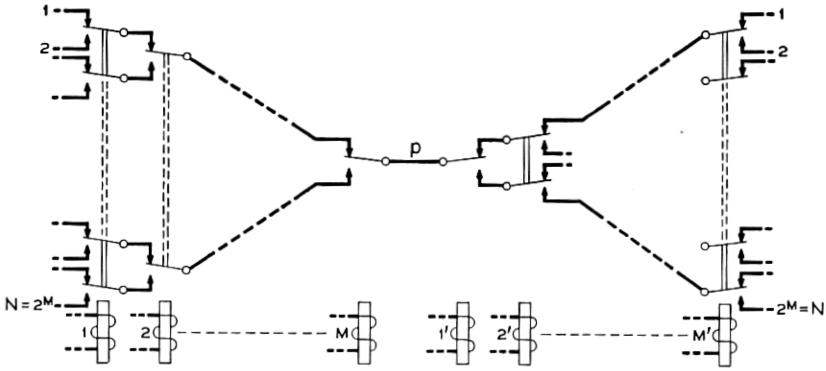

Fig. 4—Interconnecting network for Fig. 3.

from (6). The considerable discrepancy between this minimum required memory and the amount actually used in standard exchanges is due in part to the many control and supervision functions which we have ignored, and in part to statistical margins provided because of the limited access property.

The lower bound given by (6) is essentially realized with the schematic exchange of Fig. 3. Each box contains a memory $2 \log N$ and a contact network capable of interconnecting any pair of inputs, an ordered pair being associated with each possible state of the memory. Figure 4 shows such an interconnection network. By proper excitation of the memory relays 1, 2, $\cdots$, M, the point p can be connected to any of the $N = 2^m$ subscribers on the left. The relays $1'$, $2'$, $\cdots$, M' connect p to the called subscriber on the right. The general scheme of Fig. 3 is not too far from standard methods, although the contact load on the memory elements is still impractical. In actual panel, crossbar and step-by-step systems the equivalents of the memory boxes are given limited access to the lines in order to reduce the contact loads. This reduces the flexibility of interconnection, but only by a small amount on a statistical basis.

4. Relation to Information Theory

The formula $M = 2S \log N$ can be interpreted in terms of information theory.[2] When a subscriber picks up his telephone preparatory to making a call, he in effect singles out one line from the set of N, and if we regard all subscribers as equally likely to originate a call, the corresponding amount of information is $\log N$. When he dials the desired number there is a second choice from N possibilities and the total amount of information associated with the origin and destination of the call is $2 \log N$. With S possible simultaneous calls the exchange must remember $2S \log N$ units of information.

The reason we obtain the "separate memory" formula rather than the absolute minimum memory by this argument is that we have overestimated the information produced in specifying the call. Actually the originating subscribers must be one of those not already engaged, and is therefore in general a choice from less than N. Similarly the called party cannot be engaged; if the called line is busy the call cannot be set up and requires no memory of the type considered here. When these factors are taken into account the absolute minimum formula is obtained. The separate memory condition is essentially equivalent to assuming the exchange makes no use of information it already has in the form of current calls in remembering the next call.

Calculating the information on the assumption that subscribers are equally likely to originate a call, and are equally likely to call any number, corresponds to the maximum possible information or "entropy" in communication theory. If we assume instead, as is actually the case, that certain interconnections have a high *a priori* probability, with others relatively small, it is possible to make a certain statistical saving in memory.

This possibility is already exploited to a limited extent. Suppose we have two nearby communities. If a call originates in either community, the probability that the called subscriber will be in the same community is much greater than that of his being in the other. Thus, each of the exchanges can be designed to service its local traffic and a small number of intercommunity calls. This results in a saving of memory. If each exchange has N subscribers and we consider, as a limiting case, no traffic between exchanges,

[2] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 379–423, and 623–656, July and October 1948.

the total memory by (6) would be $4S \log N$, while with all $2N$ subscribers in the same exchange $4S \log 2N$ would be required.

The saving just discussed is possible because of a group effect. There are also statistics involving the calling habits of individual subscribers. A typical subscriber may make ninety per cent of his calls to a particular small number of individuals with the remaining ten per cent perhaps distributed randomly among the other subscribers. This effect can also be used to reduce memory requirements, although paper designs incorporating this feature appear too complicated to be practical.

Acknowledgment

The writer is indebted to C. A. Lovell and B. D. Holbrook for some suggestions incorporated in the paper.